Council for the Mathematical Sciences

Annex B: Bibliometric Indicators

The term 'bibliometric indicator' should not be interpreted as referring only to journal impact factors. Other bibliometric data are available and are discussed below. Over-reliance on one indicator will (a) exacerbate the effect of defects of that particular indicator as a one-dimensional measure, and (b) lead to behaviour aiming at maximising that indicator alone. It is therefore vital that if bibliometric indicators are to be employed, all available ones are brought into play, and used as judged appropriate by expert panels in the discipline.

1. Impact Factors

Thomson Scientific, formerly the Institute for Scientific Information (ISI), calculates these for all journals it covers.

1.1. Mathematics has low Impact Factors

Of greatest concern for mathematics is that Impact Factors for its journals are far lower than for other sciences, with the consequence that a mathematics journal with any hint of interdisciplinarity will have a high Impact Factor by virtue of that quality. This is easy to demonstrate. In the ISI Mathematics list (in practice Pure Mathematics), among the 181 journals in the 2005 database the complete list of those with an Impact Factor exceeding 1 is the 17 following:

Acta Mathematica,
Annals of Mathematics,
Bulletin of the American Mathematical Society,
Communications in Pure and Applied Mathematics,
Computational Complexity,
Discrete Mathematics and Theoretical Computer Science,
Discrete and Continuous Dynamical Systems,
Duke Mathematical Journal,
Geometry and Topology,
Interfaces and Free Boundaries,
Inventiones Mathematicæ,
Journal of the American Mathematical Society,
Journal of the European Mathematical Society,
Journal de Mathématiques Pures et Appliquées,
Memoirs of the American Mathematical Society,
Publications Mathématiques de l'IHES,
Random Structures and Algorithms.

No professional mathematician would concede that these are the top 17 journals in pure mathematics. The top 10 or so are included, to be sure, but others in the 17, such as *Random Structures and Algorithms*, achieve their Impact Factors by crossing subject boundaries.

Only two journals in the above list win an Impact Factor exceeding 2: *Annals of Mathematics* at 2·009, and *Journal of the American Mathematical Society* at 2·323. A comparison with the ISI Statistics and Probability list is instructive: of 81 journals, 22 have Impact Factors exceeding 1, and there are four with Impact Factors exceeding 2:

continued ...

Bioinformatics 6·019,
Biostatistics 4·529,
Econometrica 2·626,
Journal of Computational Biology 2·446.

Note that *all these have Impact Factors exceeding all pure mathematics journals.* Also, none of them are straight statistics/probability journals; they patently cross over into other fields, and achieve their high scores thereby.

This is not just a problem on the margin. One might guess that in a typical mathematics department at least half the researchers would submit papers to journals outside the core ones of their area, and which would therefore be affected by the above phenomenon. That hypothesis could be tested by analysing a small sample of departments—publicly available RAE2001 data could be used.

1.2. Impact assessed too soon for mathematics

The Impact Factor of a journal is calculated by dividing the number of current-year citations by the number of source items published in that journal during the previous two years. The time window is far too short for mathematics, where the process of refereeing and revision can easily take a year and is often longer (referees for mathematics papers will customarily check the proofs of the claims made, which takes time and usually results in need for revision). Consequently, for a published paper to be cited in a further publication within two years is a partly arbitrary outcome. The slowness of the publication process means that papers are often cited in preprint form rather than in published form—but then will not figure in any journal's Impact Factor.

The long time-period needed for impact to build up in mathematics is further evidenced by the ISI's 'Aggregate Cited Half-Life' for mathematics exceeding 10 years. This means that the majority of citations in 2005 to articles in the the ISI's Mathematics list of journals are to articles over 10 years old.

1.3. Further points regarding Impact Factors

A document on the use of Impact Factors specifically in mathematics is the *Official Statement* of the Belgian Mathematical Society, linked from the home-page of the Society `<http://bms.ulb.ac.be>`. Its main points are those above, but it also makes some further ones.

A widely cited paper on Impact Factors is T. Opthof, 'Sense and nonsense about the impact factor', *Cardiovascular Research* **33**(1997), 1–7. It is available from several web locations. The author concludes that Impact Factors are not valid for assessment of quality of individual papers, nor of individual scientists, nor of groups if they produce fewer than 100 papers in two years. The author does claim that Impact Factors are valid for assessment of journals, but his analysis is flawed by failing to consider at all the point regarding subject boundaries discussed in §1.1 above. Impact Factors are valid for assessment of journals only at best within the central core of disciplines such as mathematics.

1.4. Proprietary data and index

The Impact Factor is a proprietary product and Thomson Scientific is at liberty to change

continued ...

any aspect of it at any time. The data on which it is based are also proprietary, though derived from public information.

Thomson Scientific's *Acceptable Use Policy* for its data contains the following 'Prohibited Actions':

> "Copying/duplicating the Licensed Product or creating subsets or derivative databases from the Database, except for personal use only."
>
> "Allowing data from the Licensed Product to be made available to others or downloading the Licensed Product onto any electronic storage media or distributing or transferring the Database or Search Results in any form (printed, electronically relayed, posted to public list services or bulletin boards, or magnetically stored) to, or for the benefit, of others."

It is not obvious that mass extraction of data for use in a future RAE escapes these prohibitions. The funding councils would probably need to negotiate a contract with the Thomson Corporation.

2. Output volume

Data that quantify publication output can take the form, in decreasing order of crudeness, of

(a) number of publications,
(b) pages published,
(c) characters published.

The latter two forms are used in comparative pricing studies of journals, so data in those forms are available, presumably from Thomson Scientific's database. Use of such data as a sole method of evaluation would be rightly deplored, as indeed has reportedly been the case in Australia where page counts of publications in a list of approved journals were employed. However, data in any of the forms (a)–(c) constitute a pure, non-evaluative output measure, no less valid than the pure input measures derived from grants won. Number of outputs published is analogous to number of grants won, and numbers of pages or characters published is analogous to amount of pounds sterling won in grants. Given that such pure input measures will certainly figure in any portfolio of data that future expert panels will use, and that use of Impact Factors despite their very real problems will be hard to resist, the employment also of output data in one of the forms (a)–(c) can be only of benefit to an expert panel, and should not be dismissed *a priori*.

3. Citation counts

A count of citations to the work of an author over a chosen period is available from the Thomson Scientific database. Such data form the inputs to the calculation of Impact Factors. By taking the data over a longer period, the problem discussed in §1.2 above, of the time-frame being too short for mathematics, could be avoided. Citation counts for an author could be assembled in parallel to Impact Factor data. Practical problems of obtaining the data would, however, need to be thought through first. Given the difficulty of distinguishing scientists with the commoner name-plus-initial combinations, and the variability of the reporting of complex names, self-reporting by authors would be essential, though would then be verifiable. (Such a difficulty does not arise with Impact Factors, which are per-journal, and journals are easy to look up.)

continued ...

4. Summary

(i) Bibliometric indicators comprise several other forms of data besides the 'Impact Factor' of Thomson Scientific. Over-reliance on any one indicator is damaging, and if expert panels are to use bibliometric data, all available forms should be considered by them.

(ii) Mathematics journals have low Impact Factors. Save in the inner core of pure mathematics, and similarly for other branches of the mathematical sciences, the Impact Factor of a journal depends more on how much it extends into other subject areas, or is perceived to do so, than by its merit. Impact Factors of journals are therefore usable only if moderated by expert knowledge of the areas and sub-areas that journals fall into (or between).

(iii) The two-year citation window used in calculating the Impact Factor is too short for mathematics.

(iv) As a commercial product, the Impact Factor can by altered by its owner, Thomson Scientific, at any time. A licence for its use, or that of the data on which it is based, may have to be purchased from the Thomson Corporation.

(iv) Measures of publication quantity are available and, as pure output measures, should be used to counterbalance reliance on pure input measures such as grant volume.

(v) Citation counts, assembled with the assistance of authors but then independently verifiable, may provide information that is valuable, particularly in allowing for a time-window appropriate to mathematics.

Drafted by Charles M. Goldie, Professor of Statistics, University of Sussex.